视听时、空一致性对 Pip-and-Pop 效应的影响

唐晓雨 崔鑫忠 高敏 袁梦莹

(辽宁师范大学心理学院; 辽宁省儿童青少年健康人格评定与培养协同创新中心, 大连 116029)

摘 要 多感觉整合遵循空间原则和时间原则，有研究表明 Pip-and-Pop 效应产生的原因是多感觉整合，那么 Pip-and-Pop 效应是否同样遵循空间原则和时间原则呢？本研究采用动态视觉搜索范式，通过两个眼动实验考察空间一致性(实验 1)和时间一致性(实验 2)对 Pip-and-Pop 效应的影响。结果发现：(1)当视觉目标颜色变化伴随一个同侧的听觉刺激时所产生的 Pip-and-Pop 效应最大，对侧条件没有发现 Pip-and-Pop 效应。(2)视听刺激同时呈现所产生的 Pip-and-Pop 效应最大，随视听刺激呈现时间间隔的增加 Pip-and-Pop 效应逐渐减弱直至消失。结果表明，视听时、空一致性对 Pip-and-Pop 效应具有调节作用，这为多感觉整合在产生 Pip-and-Pop 效应中的作用提供了证据。

关键词 Pip-and-Pop 效应，视听空间一致性，视听时间一致性，动态视觉搜索范式，眼动技术

1 前言

日常生活中，我们每时每刻都在接收大量来自不同感觉通道的信息。在复杂多变的环境中寻找物体是一个非常耗时的过程。例如，在旅游景点和朋友们走散时，从人群中寻找朋友会消耗一定时间。如果朋友一边挥手一边大喊你的名字，那么找到朋友的时间会大幅缩减，寻找朋友这一过程实际上是一种伴随听觉刺激的动态视觉搜索过程。Van der Burg 等人(2008)使用动态视觉搜索范式，要求被试在多个不同倾斜角度的干扰线段中搜索垂直或水平的目标线段，搜索过程中线段颜色不断进行红绿变化。结果发现，在目标线段颜色变化的同时呈现一个纯音听觉刺激，会提高目标线段的搜索效率，这种现象被称为 Pip-and-Pop 效应(Van der Burg et al., 2008)。

有研究认为 Pip-and-Pop 效应产生的原因是多感觉整合(Chamberland et al., 2016; Van der Burg et al., 2008; Van der Burg et al., 2011)。多感觉整合(multisensory integration，MSI)是指来

收稿日期: 2021-09-03

* 辽宁省教育厅高等学校自然科学类基本科研(面上)项目(LJKZ0987)资助。

唐晓雨和崔鑫忠同为第一作者。

通信作者: 高敏, E-mail: minmin19910420@163.com

自不同感觉通道的信息相互作用，并整合为统一、连贯且有意义的知觉过程(Talsma et al., 2010; Tang et al., 2016; Van der Stoep et al., 2017)。Van der Burg 团队的研究发现，同步呈现的听觉刺激与视觉目标颜色变化后 50ms 在顶枕区发生了早期的多感觉整合，并且这一早期多感觉整合与行为学上的 Pip-and-Pop 效应显著相关，即早期多感觉整合越大，搜索效率的提高越大。Van der Burg 等人认为，通过早期整合将听觉与视觉目标绑定，使得视觉目标在众多不断变化的刺激中凸显出来，更好地捕获注意，与只有视觉目标条件相比搜索效率显著提高，所以多感觉整合在 Pip-and-Pop 效应中有着关键作用(Van der Burg et al., 2008;Van der Burg et al., 2011)。

然而，研究发现，当视觉目标不存在时听觉刺激也能提高搜索效率，所以研究者认为视听整合并不能充分地解释 Pip-and-Pop 效应。因此提出了另一种解释：即 oddball 刺激(奇异性刺激)吸引注意导致了 Pip-and-Pop 效应(Ngo & Spence, 2012; Tsai & Yeh, 2013;Zou et al., 2012)。Oddball 刺激指在其他高概率刺激中出现的低概率刺激(Pariyadath & Eagleman, 2007; Tsai & Yeh, 2013)，视觉搜索中突然出现的听觉刺激就是一个 oddball 刺激，oddball 刺激可以更有效地吸引注意(Chastain & Cheal, 1999; Escera et al., 1998; Escera et al., 2002)。研究发现当视觉目标与听觉 oddball 刺激同时呈现时，被试的主观知觉时间会延长，即被试主观上感知目标刺激的时间变长了，研究者把这种现象称为“冰冻效应”(freezing effect)(Vroomen & de Gelder, 2000;Tse et al., 2004)，无论是哪种通道的 oddball 刺激都可能导致这种效应(Ngo & Spence, 2010b;Ngo & Spence, 2012)。

Pip-and-Pop 效应产生的原因是多感觉整合还是 oddball 刺激吸引注意仍然存在争议。以往研究表明多感觉整合遵循两个原则—空间原则(the spatial rule)和时间原则(the temporal rule)，即当来自不同感觉通道的信息在大致相近的空间位置呈现时，多感觉整合效应最大(Spence, 2013; Van der Stoep et al., 2017)；当来自不同感觉通道的信息在大致接近的时间呈现时，多感觉整合效应最大(Fister et al., 2016; Stevenson et al., 2012)。所以有研究通过调控空间一致性与时间一致性来探究 Pip-and-Pop 效应产生的原因是否是多感觉整合(Van der Burg et al., 2008; Ngo & Spence, 2010a; Zou et al., 2012)。如果 Pip-and-Pop 效应遵循空间和时间原则，那么 Pip-and-Pop 效应产生的原因存在多感觉整合的作用；如果 Pip-and-Pop 效应不遵循空间和时间原则，那么 Pip-and-Pop 效应产生的原因可能是 oddball 刺激吸引注意，即突然出现的听觉刺激影响了视觉搜索效率。

一方面，前人有关视听刺激空间一致性对 Pip-and-Pop 效应影响的研究结果并不一致。Ngo 和 Spence(2010)研究发现与无效空间信息线索相比，具有空间信息的听觉刺激会提高视

觉搜索的效率，即视听刺激空间一致性影响了 Pip-and-Pop 效应。因此，Ngo 和 Spence 认为 Pip-and-Pop 效应是视听刺激整合引起的，空间位置越一致，视听刺激越容易进行整合，Pip-and-Pop 效应越大(Ngo & Spence, 2010a)。但 Fleming 等人(2020)研究表明，行为与脑电结果均显示视听刺激空间位置一致与否对 Pip-and-Pop 效应的影响差异不显著，他们由此提出空间一致性并不是 Pip-and-Pop 效应产生的必要条件，视听刺激整合并没有起到重要作用(Fleming et al., 2020)。上述两个研究结果不一致的原因可能与刺激在空间排列上的差异有关。Ngo 和 Spence(2010)将刺激随机排列在 10×10 的隐形网格中，按照搜索集大小在该隐形网格上的随机位置呈现视觉刺激。Fleming 等人(2020)虽然也如此在随机位置上呈现视觉刺激，但他们缩小了搜索区域且增加了干扰线段的方向，同时还增加了人机距离，降低了视觉上的空间关联性。所以本研究为了更好地建立空间关联性，将视觉搜索区域分为左右两侧(与中心点有明显距离)，线段整齐地排列在网格中(以搜索集 36 为例，左右分别为 3×6 的网格)，以此更明确地操纵空间一致性从而探究视听空间一致性如何影响 Pip-and-Pop 效应。

另一方面，以往有关视听刺激时间一致性对 Pip-and-Pop 效应影响的研究结果也是不一致的。Van der Burg 等人(2008)使用动态视觉搜索范式，在研究中设置了 8 种声音目标间隔(Tone-Target-Interval，TTI)条件，结果发现 TTI 为-100ms、-50ms、-25ms、0ms、25ms、100ms 时(0ms 为同时呈现，负数为听觉刺激提前呈现，正数为滞后呈现)，听觉刺激才会促进对视觉目标的搜索，视听刺激同时呈现时的促进作用最大，且声音目标间隔与反应时之间呈 U 型函数关系，即视听刺激呈现的时间一致性越高，搜索越快。Van der Burg 等人认为，同时呈现的听觉刺激以及在整合时间窗内(-100ms~100ms)呈现的听觉刺激引发的搜索效率提高是由于跨通道整合增强了视觉目标的凸显性(Van der Burg et al., 2008)。但也有研究发现视听刺激时间一致性并不会影响 Pip-and-Pop 效应，Zou 等人(2012)比较了 0ms、-100ms 以及听觉刺激随机时间呈现三种时间条件，发现这三种条件都会促进视觉目标的搜索效率，但它们之间没有显著差异，因此 Zou 等人(2012)认为 Pip-and-Pop 效应是无空间信息的听觉刺激引发的“冰冻效应”所致，即 oddball 刺激吸引注意的作用，无论听觉刺激何时呈现都会促进视觉搜索，而不是视听刺激整合作用的结果(Zou et al., 2012)。

上述研究使用的搜索集大小不同，在搜索集为 48 和 60 的研究中发现了视听时间一致性对 Pip-and-Pop 效应的影响(Van der Burg et al., 2008;Kösem & van Wassenhove, 2012)，在搜索集 36 的研究中没有发现视听时间一致性对 Pip-and-Pop 效应的影响(Zou et al., 2012)。所以搜索集大小可能在视听时间一致性影响 Pip-and-Pop 效应的过程中起着调节作用。此外，在视听空间一致性影响 Pip-and-Pop 效应的研究中，搜索集大小在其中的调节作用也是不一致的

(Ngo & Spence, 2010a;Fleming et al., 2020)。可能的原因是在搜索集较小的条件下知觉负荷较低，只要听觉刺激存在，无论其是否与视觉目标空间或时间上一致都可以提高视觉搜索效率。为了进一步确定这个问题，本研究引入搜索集大小这一变量，通过行为与眼动两方面数据探究在不同知觉负荷下视听刺激时间一致性和空间一致性对 Pip-and-Pop 效应的影响。

基于上述视听刺激空间、时间一致性对 Pip-and-Pop 效应的影响研究结论并不一致，视觉搜索任务中的行为反应时只能反映听觉刺激是否提高了搜索效率，而眼动数据可以根据眼跳和扫视等反映搜索中的实时信息处理过程。具体来说，通过平均注视次数反映知觉负荷大小，平均眼跳幅度反映在眼跳前被试获取的信息量，从而揭示听觉刺激具体如何提高搜索效率。此外，平均注视时间可以衡量是否有冰冻效应产生，从而说明 Pip-and-Pop 效应的产生是否有 oddball 刺激吸引注意的作用，进而可以探究 Pip-and-Pop 效应的产生机制。所以本研究将眼动技术运用到动态视觉搜索范式中，通过两个实验考察以下两个问题：(1)调控搜索集大小和听觉刺激呈现位置，探究视听刺激空间一致性如何影响 Pip-and-Pop 效应。(2)调控搜索集大小和听觉刺激与视觉目标的间隔时间，探究视听刺激时间一致性如何影响 Pip-and-Pop 效应。如果 Pip-and-Pop 效应产生的原因是多感觉整合，那么 Pip-and-Pop 效应遵循空间一致性和时间一致性原则，即视听刺激呈现的空间位置越一致，产生的 Pip-and-Pop 效应越大；视听刺激呈现的时间越一致，产生的 Pip-and-Pop 效应越大。如果 Pip-and-Pop 效应产生的原因是 oddball 刺激吸引注意，那么无论空间和时间一致与否，突然出现的听觉刺激都应影响视觉搜索效率。

2 实验 1：视听空间一致性对 Pip-and-Pop 效应的影响

2.1 方法

2.1.1 被试

采用 G*Power toolbox 3.1.9.2 计算实验所需最小样本量(Erdfelder et al., 2009; Faul et al., 2007)。根据听觉刺激对视觉搜索影响的研究(Ngo & Spence, 2012)，重复测量方差分析时视听空间一致性的效应量$\eta_p^2 = 0.43$，设置参数 I 类错误的概率α error prob 为 0.05，检验效能 Power (1−β err prob)为 0.9，计算得到最小样本量为 24。为避免眼动数据缺失导致样本量不足，在大学生中招募被试 32 名，其中 8 名被试在实验中有过多眨眼导致数据缺失被剔除，实际样本量为 24 人(男生 5 名，女生 19 名)，年龄 19 至 28 岁，平均年龄 22.5 岁($SD = 2.3$)。所有被试均为右利手，听力、视力或矫正视力正常，无色盲色弱。在实验完成后被试会获得

相应的报酬。

2.1.2 实验仪器和材料

实验采用 Eyelink 1000 Plus 红外反射系统眼动记录仪(SR Research，Canada)，记录被试右眼眼动数据，采样频率为 1000Hz。视觉刺激呈现在 Dell 液晶显示器上，型号为 P1914SF，屏幕可视尺寸为 19 英寸，屏幕分辨率为 1024×768 像素，刷新率为 75Hz。被试距离电脑屏幕 75cm。实验程序使用 Matlab2016a 软件和 Psychtoolbox、Eyelink Toolbox 编制。

实验 1 视觉刺激如图 1 所示，由多个干扰线段(倾斜方向分别为 22.5°，45°，67.5°，112.5°，135°，157.5°)和 1 个目标线段(水平或者垂直，目标方向在试次中平衡)组成。所有线段一半为红色(13.9cd/m²)，另一半为绿色(46.4cd/m²)，呈现在黑色(0.4cd/m²)背景的屏幕上，每个线段大小为 0.57°×0.17°。实验中有三个搜索集，分别为 36，48，60(图 1 以 36 为例)。视觉刺激分成相等的两个矩阵，呈现在屏幕左右两侧(搜索集为 36 时：每侧 3.4°×8.5°；搜索集为 48 时：每侧 5.1°×8.5°；搜索集为 60 时：每侧 6.8°×8.5°)，两侧刺激到中央注视点水平距离为 3.74°。实验过程中，线段不断进行红绿颜色变化，每次变化的线段数量不同(搜索集为 36 时，每次 1、3 或 5 个线段颜色随机变化；搜索集为 48 时，每次 1、4 或 7 个线段颜色随机变化；搜索集为 60 时，每次 1、5 或 9 个线段颜色随机变化)。变化间隔为 50、100 或者 150ms，三种变化间隔随机出现，共有九个变化间隔(一个变化周期)，以此避免视觉刺激在进行连续变化时有很高的规律性。目标线段颜色变化遵循以下 3 种规则：(1)目标线段颜色单独变化，并且每个周期只能改变一次，目标线段颜色变化的平均频率为 1.11Hz(即每 900ms 变化一次)；(2)目标颜色变化之前是 150ms 的变化间隔，目标颜色变化之后是 100ms 的变化间隔。由于视听整合的时间窗大约为±100ms，在目标变化前后的时间间隔如此设置以确保明确的视听整合；(3)每个周期的前三个间隔目标线段颜色不会变化。

听觉刺激为 1000Hz 的纯音(65 dB，60ms，包括 5ms 淡入和 5ms 淡出的时间)，通过位于屏幕后左右两侧的扬声器呈现。听觉刺激的呈现分四种情况(见图 1)：(1)无声音条件：不呈现听觉刺激；(2)同侧条件(提供有效空间信息条件)：听觉刺激与视觉目标同侧呈现，即视觉目标在右侧，听觉刺激从右声道发出；视觉目标在左侧，听觉刺激从左声道发出；(3)对侧条件(提供无效空间信息条件)：听觉刺激与视觉目标对侧呈现，即视觉目标在右侧，听觉刺激从左声道发出；视觉目标在左侧，听觉刺激从右声道发出；(4)双侧条件(不提供空间信息条件)：无论视觉刺激在左侧还是右侧，听觉刺激都从双声道发出，形成一个在中间位置的听觉刺激。

2.1.3 实验设计和程序

实验 1 采用 3(搜索集大小：36，48，60)×4(视听空间一致性：无声音，同侧，对侧，双侧)两因素被试内实验设计。在正式实验之前先进行听觉定位测试，让被试辨别听觉刺激的方向，正确区分来自不同空间位置的声音后继续进行实验。为了减少头动以获取更准确的眼动数据，实验开始前，指导被试前额抵住前额靠，下巴放在下巴托上，进行九点校准。校准完成后，开始实验。

具体实验流程如图 1 所示，每个试次呈现之前，屏幕中央会呈现一个漂移校准点，校准成功后校准点消失，然后呈现中央注视点，1000ms 后出现搜索屏，要求被试既快又准的找到目标线段，并通过按键(z 键或 m 键)反应判断目标线段是垂直还是水平，按键在被试之间平衡。如果 9 个间隔呈现完之后被试没有按键反应，则再次循环呈现实验刺激，直到被试做出反应再进入下一个试次。实验要求被试忽略听觉刺激。

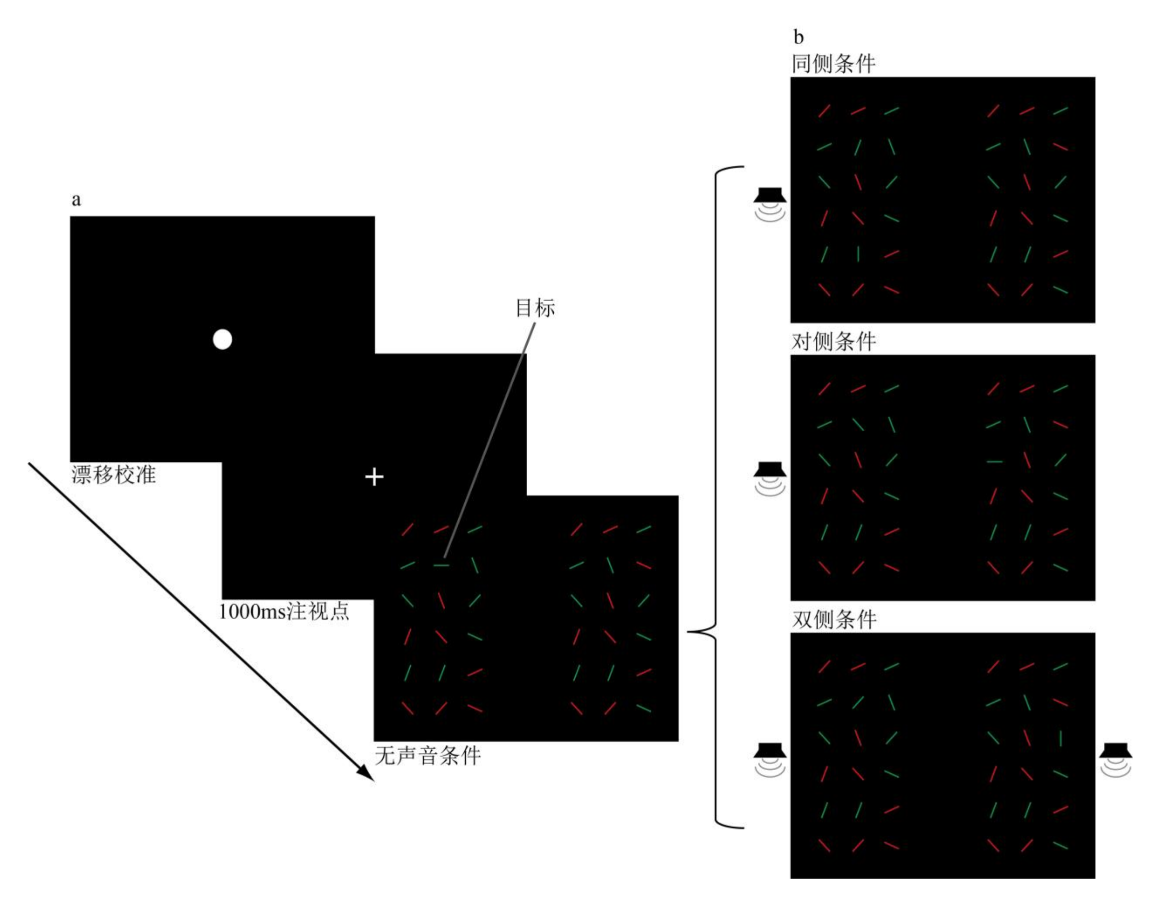


图 1　实验 1 流程图和视听刺激呈现条件图

注：图 1a 为实验 1 流程图，首先进行漂移校准，校准后呈现中央注视点，1000ms 后呈现搜索屏。搜索屏以视听刺激条件中的无声音条件为例，视觉刺激以搜索集 36 为例，包括 35 个干扰线段，1 个目标线段。图 1b 分别为视听刺激条件中的同侧条件、对侧条件、双侧条件(详情阅读方法部分)。

正式实验包括 6 组，每种搜索集 2 组，每组有 96 个试次，四种视听空间一致性条件在每组中随机呈现，共 576 个 trial。每组之前都有一个练习实验，包括 20 个试次。组之间允许被试休息，每组开始之前都要进行九点校准。整个实验大约持续 90 分钟。

2.1.4 数据分析

正确率(Accuracy)指被试正确完成试次的概率。反应时(Reaction time)指搜索屏呈现到被试做出反应之间的时间间隔。搜索效率指在正确反应的基础上，反应的快慢。反应时越长，说明被试找到目标的速度越慢，搜索效率越低；反应时越短，说明被试找到目标的速度越快，搜索效率越高。

反应时和正确率只能说明听觉刺激能否提高搜索效率，为了进一步探究听觉刺激具体如何影响搜索效率，根据以往研究选择了三个眼动指标(Zou et al., 2012)：①平均注视次数(Mean fixation number)指落入当前兴趣区注视点的次数。该指标可以有效地反映知觉负荷大小，知觉负荷越大，平均注视次数越多，搜索效率越低；②平均眼跳幅度(Mean saccade amplitude)指从当前注视到下一次注视之间的距离。眼跳幅度越大，说明被试在眼跳前获取的信息越多，搜索效率越高；③平均注视时间(Mean fixation duration)指落入当前兴趣区所有注视点的注视时间的平均。该指标反应整体的加工过程，Zou 等(2012)用该指标衡量冰冻效应，探究呈现的听觉刺激是否拓展了平均注视时间。

为了考察视听空间一致性和搜索集对 Pip-and-Pop 效应的影响，对所有指标都进行了 3(搜索集大小：36，48，60)×4(视听空间一致性：无声音，同侧，对侧，双侧)的重复测量方差分析。

2.2 结果与分析

剔除反应错误以及反应过快和过慢的试次，反应过快指目标首次变化前被试就已经做出反应的试次，反应过慢指被试在目标变化了 10 次及以上依旧没有做出反应的试次(Van der Burg et al., 2008)。最终剔除试次占总试次的 5.5%。

2.2.1 正确率

所有被试的正确率均达到 98%以上，正确率方差分析结果发现搜索集大小主效应不显著，$F(2，46) = 0.74$，$p = 0.48$；视听刺激空间一致性主效应不显著，$F(3，69) = 0.25$，$p = 0.86$；二者交互作用不显著，$F(6，138) = 0.66$，$p = 0.68$。各个实验条件下的正确率均较高。

2.2.2 反应时

反应时方差分析结果发现搜索集大小主效应显著，$F(2，46) = 5.34$，$p = 0.008$，$\eta_p^2 = 0.19$，进一步分析表明，搜索集为 60 条件的反应时(3.5s)显著长于搜索集为 48 条件(3.3s，$p = 0.03$)

和搜索集为 36 条件的反应时(3.2s，$p = 0.004$)，后两者之间无显著差异($p = 0.47$)。视听刺激空间一致性主效应显著，$F(3，69) = 11.86$，$p < 0.001$，$\eta_p^2 = 0.34$，进一步分析表明，同侧条件的反应时(3.07s)最短，并且显著短于双侧条件(3.27s，$p = 0.006$)、对侧条件(3.50s，$p < 0.001$)以及无声音条件(3.43s，$p < 0.001$)的反应时。双侧条件的反应时也显著短于对侧条件($p = 0.005$)和无声音条件($p = 0.038$)的反应时。对侧条件的反应时最长，并且和无声音条件无显著差异($p = 0.35$)。搜索集大小和视听空间一致性的交互作用不显著，$F(6，138) = 0.46$，$p = 0.84$(见图 2a)。

2.2.3 眼动数据

（1）平均注视次数

平均注视次数方差分析结果发现搜索集大小主效应显著，$F(2，46) = 3.44$，$p = 0.04$，$\eta_p^2 = 0.13$，进一步分析表明，搜索集为 60 条件的注视次数(14.20 次)显著多于搜索集为 36 条件(13.11 次，$p = 0.01$)的注视次数。搜索集为 48 条件(13.57 次)的注视次数与搜索集为 60 条件($p = 0.183$)和搜索集为 36 条件($p = 0.261$)的注视次数都无显著差异。视听刺激空间一致性主效应显著，$F(3，69) = 7.45$，$p < 0.001$，$\eta_p^2 = 0.25$，进一步分析表明，同侧条件的注视次数(12.77 次)最少，并且显著少于双侧条件(13.45 次，$p = 0.02$)、对侧条件(14.16 次，$p = 0.002$)以及无声音条件(14.12 次，$p = 0.003$)的注视次数。双侧条件的注视次数显著少于对侧条件($p = 0.036$)以及无声音条件($p = 0.048$)的注视次数。对侧注视次数最多，并且和无声音条件无显著差异($p = 0.90$)。搜索集大小和视听空间一致性的交互作用不显著，$F(6，138) = 0.92$，$p = 0.48$(见图 2b)。

（2）平均眼跳幅度

平均眼跳幅度方差分析结果发现搜索集大小主效应显著，$F(2，46) = 63.31$，$p < 0.001$，$\eta_p^2 = 0.73$，进一步分析表明，搜索集为 60 条件的眼跳幅度(4.96°)显著大于搜索集为 48 条件(4.50°，$p < 0.001$)和搜索集为 36 条件(4.21°，$p < 0.001$)的眼跳幅度，搜索集为 48 条件的眼跳幅度显著大于搜索集为 36 条件($p < 0.001$)的眼跳幅度。视听空间一致性主效应显著，$F(3，69) = 5.63$，$p = 0.002$，$\eta_p^2 = 0.20$，进一步分析表明，同侧条件的眼跳幅度(4.64°)最大，且显著大于双侧条件(4.55°，$p = 0.017$)，对侧条件(4.54°，$p = 0.028$)以及无声音条件(4.49°，$p < 0.001$)的眼跳幅度，双侧条件与对侧条件($p = 0.875$)和无声音条件($p = 0.154$)无显著差异，对侧条件与无声音条件也无显著差异($p = 0.156$)。

搜索集大小和视听空间一致性的交互作用显著，$F(6，138) = 5.61$，$p < 0.001$，$\eta_p^2 = 0.20$。简单效应分析发现，只有在搜索集为 60 条件下，不同视听刺激空间一致性条件有显著差异，

$F(3, 21) = 13.05$，$p < 0.001$，$\eta_p^2 = 0.65$。同侧条件的眼跳幅度(5.14°)最大，并且显著大于对侧条件(4.87°，$p < 0.001$)以及无声音条件(4.80°，$p = 0.001$)的眼跳幅度，与双侧条件(5.03°，$p = 0.34$)的眼跳幅度无显著差异。双侧条件的眼跳幅度也显著大于无声音条件($p = 0.004$)和对侧条件($p = 0.018$)的眼跳幅度。无声音条件的眼跳幅度最小，和对侧条件无显著差异($p = 0.42$)。在搜索集为 48 和搜索集为 36 条件下，不同视听刺激空间一致性条件无显著差异，$F(3, 21) = 2.37$，$p = 0.099$；$F(3, 21) = 1.24$，$p = 0.32$(见图 2c)。

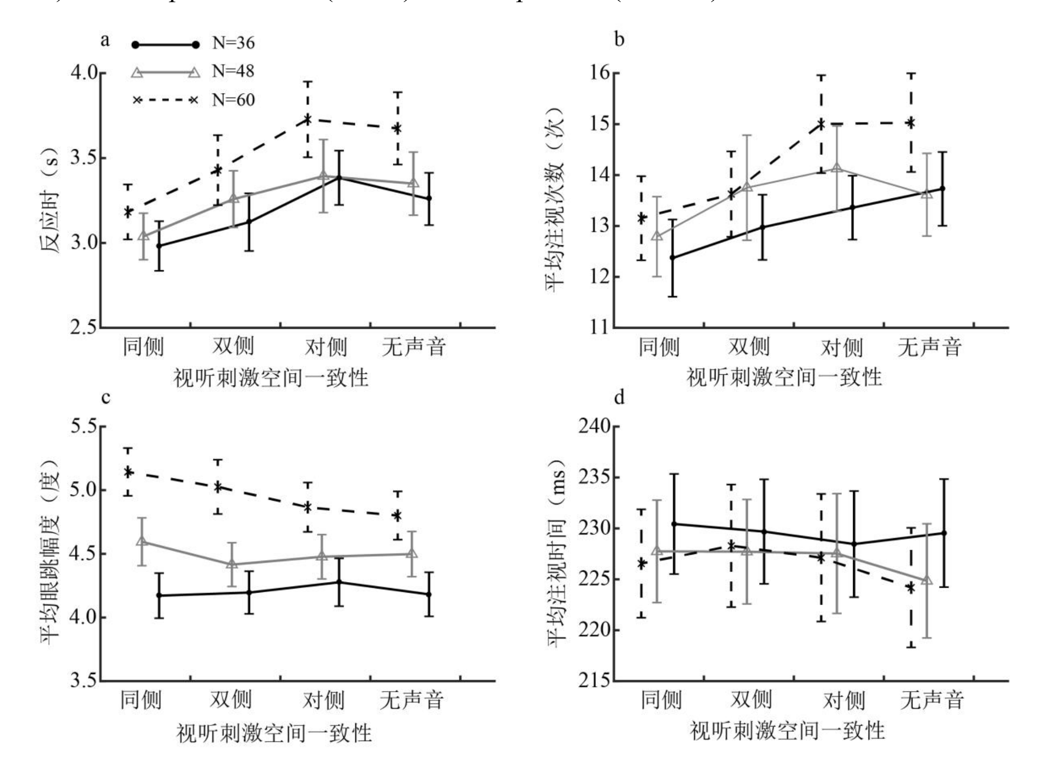


图 2　实验 1 结果图

注：图 a、b、c、d 分别为不同条件下的反应时(s)、平均注视次数(次)、平均眼跳幅度(度)、平均注视时间(ms)的结果。

（3）平均注视时间

平均注视时间方差分析结果发现搜索集大小主效应不显著，$F(2, 46) = 0.51$，$p = 0.60$；视听刺激空间一致性主效应不显著，$F(3, 69) = 1.52$，$p = 0.22$；二者交互作用不显著，$F(6, 138) = 0.87$，$p = 0.52$。(见图 2d)。

本实验进一步分析了四种条件下(视觉目标出现在同一位置上)的眼动轨迹和热点图(以搜索集 60 为例，见图 3)。与无声音的基线条件相比，同侧条件下的搜索效率最高，注视次数最少，双侧条件次之。虽然对侧条件的注视次数比无声音的注视次数少一些，但相比于其他两个条件，注视次数明显增多，搜索效率明显降低。

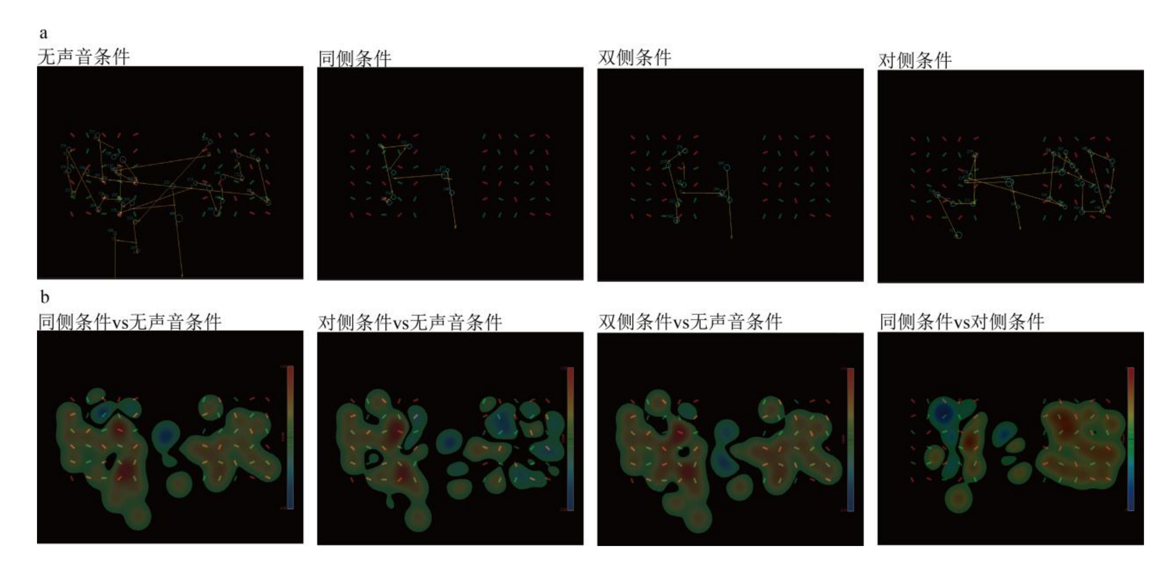


图 3 实验 1 眼动轨迹和热点图

注：图 3a 分别为四种声音条件的眼动轨迹图，其中黄色线路代表眼动轨迹，小圆圈代表注视次数。图 3b 为各个条件之间对比的差异热点图，其中每个热点图中 vs 左侧条件为蓝色，右侧条件为红色。

2.3 讨论

实验 1 通过操纵搜索集大小与听觉刺激呈现位置，考察视听空间一致性对 Pip-and-Pop 效应的影响。结果发现，与无声音的基线条件相比，同侧条件下的反应时最短，搜索过程中的平均注视次数最少，平均眼跳幅度最大，表明同侧条件下的搜索效率最高，即同侧条件所产生的 Pip-and-Pop 效应最大；其次双侧条件也产生了 Pip-and-Pop 效应，但产生的效应比同侧条件弱；而对侧条件在上述指标中与无声音条件无显著差异，即对侧条件下没有发现 Pip-and-Pop 效应。实验结果表明，视听刺激空间一致性对 Pip-and-Pop 效应产生了影响，视听刺激在空间位置上越一致，所产生的 Pip-and-Pop 效应越大。

同侧条件下，当听觉刺激从目标侧声道发出，在听觉刺激的有效空间信息的作用下，被试可以直接定位到目标侧进行搜索，在目标侧通过少量注视，大幅度眼跳快速地搜索到目标线段，搜索时间短，效率高，所以产生的 Pip-and-Pop 效应最大。这说明听觉刺激可能被当作提示线索，从而导致搜索效率的提高。但听觉刺激引导被试到同侧后并没有立即定位到目标位置，在接下来的搜索过程中，听觉刺激仍然起到增强视觉目标变化凸显性的作用，因为在双侧条件下，即使没有提示线索的作用，也发现了 Pip-and-Pop 效应。所以可能是听觉刺激的提示及其与视觉目标整合的共同作用导致搜索效率的提高，产生 Pip-and-Pop 效应。而对侧条件下，当听觉刺激从非目标侧声道发出，在听觉刺激的无效空间信息作用下，被试先在非目标侧进行大量搜索后再回到目标侧搜索，在非目标侧进行多次注视，眼跳幅度不断缩小进行细致搜索，没有发现目标线段后再回到目标侧进行搜索，相比于同侧条件，对侧条件在非目标侧的无效搜索产生了搜索成本，通过两个条件的注视次数差异热点图也可以明显地

发现对侧条件在非目标侧上的搜索成本(见图 3)，所以同侧条件的搜索效率显著高于对侧条件。以往研究发现与不提供空间信息的听觉刺激相比，提供空间信息的听觉刺激能够减少被试辨别和检测目标的潜伏期(Perrott et al., 1990)。因此，听觉线索和视觉目标在空间位置上一致可以有效的将被试的空间注意指向视觉目标，更加利于被试搜索到视觉目标。此外，也有研究发现来自不同通道的刺激位置接近时，产生的多感觉相互作用最强；不同通道的刺激之间空间距离增加时多感觉相互作用减弱(Lewald et al., 2001; Slutsky & Recanzone, 2001; Soto-Faraco et al., 2003; Stein & Stanford, 2008)。综上，本实验中的同侧条件可能是听觉刺激的提示作用和视、听刺激方位一致产生了较强的多感觉相互作用，二者共同促进了搜索效率，所以产生的 Pip-and-Pop 效应最大；而对侧条件可能是听觉刺激的无效提示作用和两个通道刺激方位不一致，距离增大，多感觉相互作用减弱，所以不足以产生 Pip-and-Pop 效应。

实验中的双侧条件也产生了 Pip-and-Pop 效应，因为当听觉刺激从双声道发出，形成了无空间信息的听觉刺激，无空间信息的听觉刺激和视觉目标同步呈现时，听觉刺激促进了视觉目标的识别(Giard & Peronnet, 1999; Ngo & Spence, 2012; Yang et al., 2014)。研究中对目标识别的促进可能是多感觉整合的促进效应(multisensory performance improvement effects)，类似于以往研究中的冗余信号效应(redundant signals effect, RSE)(Mishler & Neider, 2016; Mishler & Neider, 2018)，即个体对多通道刺激的反应比单通道刺激反应更加快速和准确(van den Brink et al., 2014; Van der Stoep et al., 2015)。以往研究也发现多感觉整合后的刺激比单通道刺激更容易吸引注意(Lunn et al., 2019)。此外，Zou 等(2012)研究发现，当同步呈现听觉刺激时，与非目标侧相比，目标侧的注视次数更多，说明被试更可能拒绝没有视觉目标的一侧，并且快速的将眼跳指向目标所在的一侧(Zou et al., 2012)。虽然双侧条件下的听觉刺激没有提供有效空间信息，但双通道的整合促进作用和双通道刺激吸引更多注意使搜索效率提高，无空间信息的听觉刺激对于眼动的调节也可以使搜索效率提高，从而产生 Pip-and-Pop 效应。

综上所述，Pip-and-Pop 效应是通过减少注视次数，增加眼跳幅度，进而缩短搜索时间产生的。实验 1 无论是反应时结果还是眼动指标结果都表明，视听刺激空间一致性对 Pip-and-Pop 效应产生了影响，视听刺激在空间位置上越一致，所产生的 Pip-and-Pop 效应越大。

3 实验 2：视听时间一致性对 Pip-and-Pop 效应的影响

实验 2 的目的是考察 Pip-and-Pop 效应是否遵循时间原则，即时间一致性对 Pip-and-Pop 效应是否具有调节作用。

3.1 方法

3.1.1 被试

采用 G*Power toolbox 3.1.9.2 计算实验所需最小样本量(Erdfelder et al., 2009; Faul et al., 2007)。根据听觉刺激对视觉搜索影响的研究(Van der Burg et al., 2008)，重复测量方差分析时视听时间一致性的效应量$\eta_p^2 = 0.53$，设置参数 I 类错误的概率α error prob 为 0.05，检验效能 Power (1−β err prob)为 0.9，计算得到最小样本量为 24。为避免眼动数据缺失导致样本量不足，在大学生中招募被试 35 名，其中 8 名被试在实验中有过多眨眼导致数据缺失被剔除，实际样本量为 27 人(男生 5 名，女生 22 名)，年龄 18 至 25 岁，平均年龄 20.7(*SD* = 2.4)。所有被试均为右利手，听力、视力或矫正视力正常，无色盲色弱。在实验完成后被试会获得相应的报酬。

3. 1. 2 实验仪器和材料

实验仪器同实验 1。视觉刺激材料有两点不同于实验 1：第一，本实验中搜索集大小为 36 和 60，由于实验 1 中只在平均眼跳幅度这个指标上发现了搜索集 48 与 36 和 60 有显著差异，而反应时、平均注视次数以及平均注视时间上均无显著差异，所以取消了搜索集为 48 的条件。第二，在听觉刺激与视觉目标变化不一致呈现条件下，为了使听觉刺激与视觉干扰刺激的变化同步呈现，将本实验中的线段变化间隔由实验 1 中的 50ms、100ms、150ms 三种改为 100ms 一种变化间隔。

听觉刺激材料同实验 1。听觉刺激与视觉目标颜色变化之间的时间间隔(TTI：tone – target interval)有 6 种：(1)无声音条件：不呈现听觉刺激；(2)TTI = -200ms 条件：听觉刺激先于视觉目标颜色变化 200ms 呈现；(3)TTI = -100ms 条件：听觉刺激先于视觉目标颜色变化 100ms 呈现；(4)TTI = 0ms 条件：听觉刺激与视觉目标颜色变化同步呈现；(5)TTI = 100ms 条件：听觉刺激晚于视觉目标颜色变化 100ms 呈现；(6)TTI = 200ms 条件：听觉刺激晚于视觉目标颜色变化 200ms 呈现。

3.1.3 实验设计和程序

实验采用 2(搜索集大小：36，60)×6(视听时间一致性：无声音条件，-200ms，-100ms，0ms，100ms，200ms)两因素被试内实验设计。实验具体流程如图 4a 所示，与实验 1 的实验

流程完全一致。正式实验包括 12 组，每种搜索集大小和视听刺激时间一致性条件分别在不同组中呈现，每组有 48 个试次，共 576 个试次。每组之前都有 12 个练习试次。组间允许被试休息，整个实验大约持续 90 分钟。

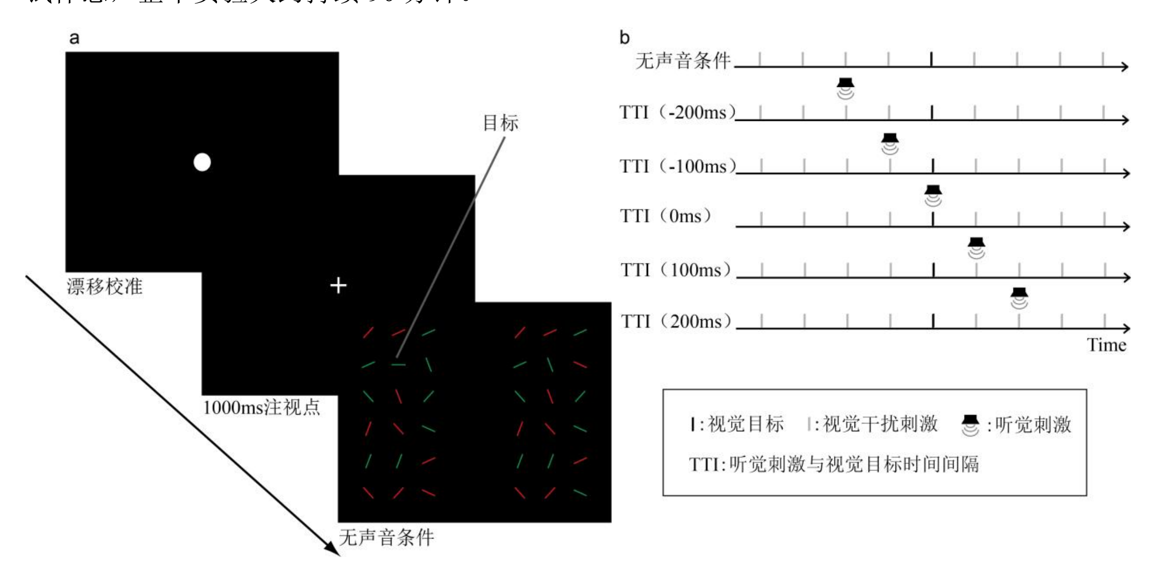


图 4　实验 2 流程和视听刺激条件图

注：图 4a 为实验 2 流程图，首先进行漂移校准，校准后呈现中央注视点，1000ms 后呈现搜索屏。搜索屏以视听刺激条件中的无声音条件为例，视觉刺激以搜索集 36 为例，包括 35 个干扰线段，1 个目标线段。图 4b 是 6 种 TTI(听觉刺激与视觉刺激之间的时间间隔：tone-target-interval)条件，其中黑色线段是指目标线段颜色变化；灰色线段是指干扰线段颜色变化；小喇叭是指发出听觉刺激(详情阅读方法部分)。

3.1.4 数据分析

选取指标同实验 1。为了考察视听时间一致性和搜索集大小对 Pip-and-Pop 效应的影响，对所有指标都进行了 2(搜索集大小：36，60)×6(TTI：无声音，-200ms，-100ms，0ms，100ms，200ms)的重复测量方差分析。

3.2 结果与分析

数据剔除标准同实验 1。最终剔除试次占总试次的 2%。

3.2.1 正确率

所有被试正确率均达到 98%以上，正确率方差分析结果发现搜索集大小主效应不显著，$F(1, 26) = 0.49$，$p = 0.49$；TTI 主效应不显著，$F(5, 130) = 0.89$，$p = 0.49$；二者交互作用不显著，$F(5, 130) = 1.64$，$p = 0.15$。各个实验条件下的正确率均较高。

3.2.2 反应时

反应时方差分析结果发现搜索集大小主效应显著，$F(1, 26) = 102.81$，$p < 0.001$，$\eta_p^2 = 0.80$，进一步分析表明，搜索集为 60 条件的反应时(4.61s)显著长于搜索集为 36 条件(2.82s)的反应时；TTI 主效应显著，$F(5, 130) = 8.91$，$p < 0.001$，$\eta_p^2 = 0.26$，进一步分析表明，TTI 为 0ms(2.73s)时，反应时最短，并显著短于其他 5 种 TTI 条件($ps < 0.001$)。其次是-100ms 条件

(3.26s，p = 0.001)和 100ms 条件(3.44s，p = 0.021)的反应时都显著短于无声音条件(4.40s)的反应时；-200ms 条件(3.85s，p = 0.174)和 200ms 条件的反应时(4.60s，p = 0.744)与无声音条件的反应时无显著差异；-200ms 条件的反应时显著长于-100ms 条件(p = 0.001)的反应时，200ms 条件的反应时显著长于 100ms 条件(p = 0.001)的反应时。

搜索集大小与 TTI 的交互作用显著，$F(5, 130) = 3.96$，$p = 0.002$，$\eta_p^2 = 0.13$，简单效应分析发现，两种搜索集条件下 TTI 主效应显著：在搜索集为 60 条件下，$F(5, 22) = 17.26$，$p < 0.001$，$\eta_p^2 = 0.80$，TTI 为 0ms(3.35s)时，反应时最短，并显著短于其他 5 种 TTI 条件(ps < 0.001)。其次是-100ms 条件(3.96s，$p = 0.002$)和 100ms 条件(4.22s，$p = 0.02$)的反应时都显著短于无声音条件(5.65s)的反应时；-200ms 条件(4.77s，$p = 0.15$)和 200ms 条件(5.74s，p=0.921)的反应时与无声音条件的反应时无显著差异。-200ms 条件的反应时显著长于-100ms 条件(p = 0.001)的反应时，200ms 条件的反应时显著长于 100ms 条件($p < 0.001$)的反应时。在搜索集为 36 条件下，$F(5, 22) = 8.30$，$p < 0.001$，$\eta_p^2 = 0.65$，0ms 条件的反应时(2.11s)也最短，并显著短于其他 5 种 TTI 条件($ps < 0.001$)。其次是-100ms 条件(2.57s)的反应时显著短于无声音条件(3.15s，$p = 0.01$)，100ms 条件(2.67s)的反应时与无声音条件无显著差异($p = 0.092$)；-200ms 条件(2.93s，$p = 0.372$)和 200ms 条件(3.47s，$p = 0.489$)的反应时与无声音条件的反应时无显著差异。-200ms 条件的反应时与-100ms 条件的反应时差异显著($p = 0.018$)，200ms 条件的反应时与 100ms 条件的反应时差异显著($p = 0.013$)(见图 5a)。

3.2.3 眼动数据

（1）平均注视次数

平均注视次数方差分析结果发现搜索集大小主效应显著，$F(1, 26) = 113.85$，$p < 0.001$，$\eta_p^2 = 0.81$，进一步分析表明，搜索集为 60 条件的注视次数(16.88 次)显著多于搜索集为 36 条件(10.31 次)的注视次数。TTI 主效应显著，$F(5, 130) = 11.47$，$p < 0.001$，$\eta_p^2 = 0.31$，进一步分析表明，TTI 为 0ms(9.85 次)时，注视次数最少，并显著少于其他 5 种 TTI 条件(ps < 0.001)。其次是-100ms 条件(11.93 次，$p < 0.001$)和 100ms 条件(12.65 次，$p = 0.006$)的注视次数显著少于无声音条件(16.46 次)的注视次数。-200ms 条件(14.03 次，$p = 0.082$)和 200ms 条件(16.64 次，$p = 0.926$)的注视次数与无声音条件的注视次数无显著差异；-200ms 条件的注视次数显著多于-100ms 条件的注视次数($p < 0.001$)，200ms 条件的注视次数显著多于 100ms 条件的注视次数($p < 0.001$)。

搜索集大小与 TTI 的交互作用显著，$F(5, 130) = 5.11$，$p < 0.001$，$\eta_p^2 = 0.16$，简单效应分析发现，两种搜索集条件下 TTI 主效应显著：在搜索集为 60 条件下，$F(5, 22) = 19.33$，

$p < 0.001$，$\eta_p^2 = 0.82$，TTI 为 0ms(12.02 次)时，注视次数最少，并显著少于其他 5 种 TTI 条件($ps < 0.001$)。其次是-100ms 条件(14.43 次，$p = 0.001$)和 100ms 条件(15.56 次，$p = 0.01$)的注视次数显著少于无声音条件(20.90 次)的注视次数；-200ms 条件(17.27 次，$p = 0.079$)和 200ms 条件(20.10 次，$p = 0.943$)的注视次数与无声音条件的注视次数无显著差异；-200ms 条件的注视次数显著多于-100ms 条件的注视次数($p < 0.001$)，200ms 条件的注视次数显著多于 100ms 条件的注视次数($p < 0.001$)。在搜索集为 36 条件下，$F(5，22) = 10.19$，$p < 0.001$，$\eta_p^2 = 0.70$，0ms 条件的注视次数(7.68 次)也最少，并显著少于其他 5 种 TTI 条件($ps < 0.001$)。其次是-100ms 条件(9.44 次，$p = 0.002$)和 100ms 条件(9.75 次，$p = 0.019$)的注视次数显著少于无声音条件(12.02 次)的注视次数；-200ms 条件(10.79 次，p=0.18)和 200ms 条件(12.19 次，$p = 0.902$)的注视次数与无声音条件的注视次数无显著差异。-200ms 条件的注视次数显著多于-100ms 条件的注视次数($p = 0.018$)，200ms 条件的注视次数显著多于 100ms 条件的注视次数($p = 0.006$)(见图 5b)。

（2）平均眼跳幅度

平均眼跳幅度方差分析结果发现搜索集大小主效应显著，$F(1，26) = 9.01$，$p = 0.006$，$\eta_p^2 = 0.26$，进一步分析表明，搜索集为 60 条件的眼跳幅度(4.59°)显著大于搜索集为 36 条件(4.37°)的眼跳幅度。TTI 主效应显著，$F(5，130) = 4.67$，$p = 0.001$，$\eta_p^2 = 0.15$，进一步分析表明，TTI 为 0ms(4.78°)时，眼跳幅度最大，并显著大于其他 5 种 TTI 条件($ps < 0.021$)。其次是-100ms 条件(4.49°，$p = 0.032$)和 100ms 条件(4.50°，$p = 0.003$)显著大于无声音条件(4.25°)的眼跳幅度；-200ms 条件(4.43°，$p = 0.055$)和 200ms(4.45°，$p = 0.151$)的眼跳幅度与无声音条件的眼跳幅度无显著差异；-200ms 条件的眼跳幅度与-100ms 条件的眼跳幅度无显著差异($p = 0.545$)，200ms 条件的眼跳幅度与 100ms 条件的眼跳幅度无显著差异($p = 0.727$)。搜索集大小与 TTI 的交互作用不显著，$F(5，130) = 2.13$，$p = 0.066$(见图 5c)。

（3）平均注视时间

平均注视时间方差分析结果发现 TTI 的主效应显著，$F(5，130) = 4.39$，$p = 0.001$，$\eta_p^2 = 0.14$，进一步分析表明，TTI 为 0ms(245.04ms)时，注视时间最长，并显著长于其他 5 种 TTI 条件($ps < 0.026$)。100ms 条件(237.03ms，$p = 0.031$)和-100ms 条件(236.11ms，$p = 0.036$)的注视时间显著长于无声音条件(228.14ms)的注视时间；200ms 条件(237.25ms，$p = 0.061$)和-200ms 条件(234.59ms，$p = 0.187$)的注视时间与无声音条件的注视时间无显著差异；200ms 与 100ms 的注视时间无显著差异($p = 0.909$)，-200ms 与-100ms 的注视时间无显著差异($p = 0.691$)。搜索集大小主效应不显著，$F(1，26) = 1.42$，$p = 0.25$；搜索集大小与 TTI 的交互作

用不显著，$F(5, 130) = 1.26$，$p = 0.29$(见图 5d)。

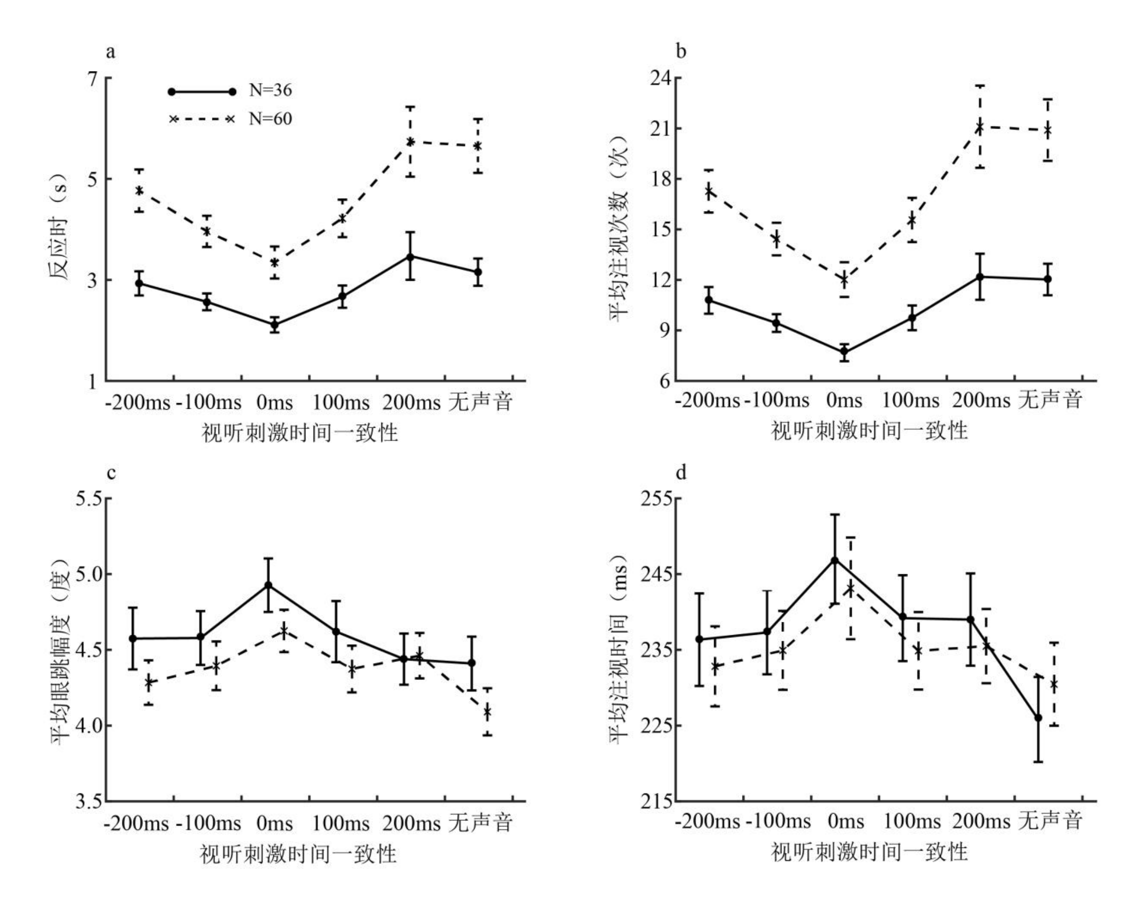


图 5 实验 2 结果图

注：图 a、b、c、d 分别为不同条件下的反应时(s)、平均注视次数(次)、平均眼跳幅度(度)、平均注视时间(ms)的结果。

3.3 讨论

实验 2 通过操纵搜索集大小和听觉刺激与视觉目标的间隔时间，考察视听刺激时间一致性对 Pip-and-Pop 效应的影响。结果发现，与无声音的基线条件相比，听觉刺激与视觉目标颜色变化之间的时间间隔(TTI)在 0ms，-100ms，100ms 条件下，搜索反应时都显著变短，搜索过程中的注视次数减少，平均眼跳幅度增加，平均注视时间变长；当 TTI 为-200ms 和 200ms 时，与无声音条件无显著差异。这说明 TTI 在-100ms 到 100ms 产生了 Pip-and-Pop 效应，其中 TTI 为 0ms 时 Pip-and-Pop 效应最大，-200ms 和 200ms 两种条件没有产生 Pip-and-Pop 效应。实验结果表明视听刺激时间越一致，Pip-and-Pop 效应越大。

研究结果表明，反应时与 TTI 之间的关系呈 U 型曲线：TTI 的绝对值越小，反应时越短。TTI 在-100ms 到 100ms 产生了 Pip-and-Pop 效应，这与视听整合的时间窗口一致(Lewald et al., 2001; Lewald & Guski, 2003)。并且在早期研究中发现，与视觉目标时间上同步的听觉刺激会增加视觉刺激的感知强度(Stein et al., 1996)，所以 0ms 条件下同步的听觉刺激可能会增强对视觉目标的感知，产生的 Pip-and-Pop 效应最大。而-200ms 和 200ms 条件不在视听

整合的时间窗范围内，即使听觉刺激存在，但不能有效的与视觉目标进行整合，所以没有产生 Pip-and-Pop 效应。

在平均注视次数眼动指标中发现了与反应时相同的结果，平均注视次数与 TTI 之间的关系呈 U 型曲线：TTI 的绝对值越小，平均注视次数越少；平均眼跳幅度与平均注视时间结果也表明只有 TTI 在 0ms，-100ms，100ms 条件下，才会优化搜索的眼动行为。这种在眼动行为上的优化在以往研究中被认为是冰冻效应(freezing effect)所致，Zou 等(2012)使用平均注视时间来量化冰冻效应。即无空间信息的听觉刺激“冰冻”了眼动，突然出现的听觉刺激抑制了眼跳，使眼跳延迟，从而搜索过程中的平均注视时间变长，那么被试就可以在更长的时间和更广阔的空间内通过增大眼跳幅度、减少注视次数对视觉目标进行搜索，从而更精确地引导接下来的注视指向目标，加快对目标的定位。此外，由这种“冰冻”所引起的时间变长不仅可以提高目标在当前区域的搜索效率，也可以快速拒绝目标不在的区域，从而提高整体的搜索效率(Zou et al., 2012)。与以往研究结果不同的是，本研究发现 0ms 条件时的平均注视时间最长，这说明听觉刺激对视觉搜索的促进作用并不全是冰冻效应引起的。综上表明 Pip-and-Pop 效应遵循时间一致性原则，视听刺激呈现时间越一致，产生的 Pip-and-Pop 效应越大。

4 总讨论

本研究采用动态视觉搜索范式，通过两个实验考察视听刺激空间一致性、时间一致性对 Pip-and-Pop 效应的影响。实验 1 发现同侧条件所产生的 Pip-and-Pop 效应最大，双侧条件也产生了 Pip-and-Pop 效应，对侧条件没有产生 Pip-and-Pop 效应；实验 2 发现视听刺激同时呈现所产生的 Pip-and-Pop 效应最大，随视听刺激呈现时间间隔的增加 Pip-and-Pop 效应减弱。研究表明，视听刺激在空间、时间上越一致，对 Pip-and-Pop 效应越大。

4.1 知觉负荷在视听时、空一致性对 Pip-and-Pop 效应影响中的调节作用

本研究的两个实验都发现了搜索集大小对 Pip-and-Pop 效应的影响，实验 1 在平均眼跳幅度指标中发现了搜索集大小和视听空间一致性存在交互作用；实验 2 在反应时和注视次数两个指标中发现了搜索集大小和视听时间一致性存在交互作用。前人研究表明，在识别单个目标时增加干扰刺激数量，意味着存在更大的认知负担(Pluta et al., 2011)，那么研究中增大搜索集也就意味着增大了知觉负荷，两个实验结果都表明搜索集大的条件所产生的知觉负荷和 Pip-and-Pop 效应大。

实验 1 发现只有在搜索集为 60 条件下，视听空间一致性才会增加平均眼跳幅度。根据知觉负荷理论，与无负荷或低负荷条件相比，在高知觉负荷条件下被试需要集中所有注意资源参与到任务中(Lavie, 2005)，本研究中相比于其他两个搜索集条件，搜索集为 60 条件下被试投入了更多的注意资源，所以产生的 Pip-and-Pop 效应最大。

实验 2 发现在搜索集为 36 条件下反应时和注视次数两个指标的听觉刺激作用均有减小。在反应时和注视次数这两个指标上，搜索集为 36 条件下仅发现视听刺激间隔在 0ms 条件和 -100ms 条件时对搜索有促进作用，100ms 条件对搜索不再有促进作用。此外，在 Zou 等(2012)使用搜索集为 36 的研究中表明，反应时与眼动结果都发现无论视听刺激呈现时间是否一致，对搜索都有促进作用，但时间条件之间没有显著差异(Zou et al., 2012)。其原因可能是所使用的搜索集较小，相应的知觉负荷也小。已有研究表明知觉负荷影响听觉刺激对视觉搜索的促进作用，在搜索复杂而非简单的视觉刺激时，声音的促进效果更强(Knoeferle et al., 2016)，那么在搜索集为 36 条件下，视觉搜索任务简单，仅根据视觉目标的颜色变化也能快速搜索到视觉目标，而在搜索集为 60 条件中，只有同步的听觉刺激与视觉目标整合吸引更多的注意，才能更快的搜索到视觉目标。综上两个实验结果都说明知觉负荷在视听空间一致性和视听时间一致性影响 Pip-and-Pop 效应中起到了调节作用。

4.2 Pip-and-Pop 效应的产生原因

本研究表明视听刺激空间一致性和时间一致性影响 Pip-and-Pop 效应，这说明 Pip-and-Pop 效应遵循多感觉整合的两个重要原则，即实验 1 研究结果符合多感觉整合的空间原则；实验 2 研究结果符合多感觉整合的时间原则。基于以往研究和本研究的结果可以推测 Pip-and-Pop 效应产生的原因至少有一部分是由多感觉整合所引起的。以往研究中对 Pip-and-Pop 效应的解释除了多感觉整合以外，也有研究者认为是 oddball 刺激吸引注意产生了 Pip-and-Pop 效应。Ngo 等(2010)研究发现，呈现视觉线索时对视觉目标的促进作用与呈现听觉线索时对视觉目标的促进作用没有显著差异，这表明无论刺激是单通道还是双通道，只要与视觉目标同时呈现的线索使视觉目标变成一个 oddball 刺激，这种促进效应就能出现(Ngo & Spence, 2010b)。

那么本研究中所发现的 Pip-and-Pop 效应是否也可以用 oddball 刺激吸引注意来解释呢？第一，本研究中除了基线的无声音条件外，两个实验所有的条件都是通过视听双通道刺激呈现的，所以即使本研究中与视觉目标同时呈现的听觉线索使视觉目标变成 oddball 刺激吸引注意可以产生 Pip-and-Pop 效应，但也无法排除视听整合在其中的作用。第二，本研究中的听觉线索包含了有空间信息和无空间信息听觉线索两种，以往发现 oddball 刺激吸引注意的

视觉搜索研究中只有无空间信息听觉线索，二者情况并不一致。第三，本研究中所使用的动态视觉搜索范式与以往发现 oddball 刺激吸引注意的视觉搜索范式不同。例如，本研究中被试在动态和杂乱的干扰物中搜索视觉目标，同步的听觉线索可以让视觉目标比干扰物更显著，也就是说，让视觉目标从周围的干扰物刺激中“pop-out”，从而提高搜索效率。而发现 oddball 刺激吸引注意的研究使用的是序列呈现的视觉搜索范式，视觉目标的呈现时间非常短暂，同步的线索以一种冰冻的方式，允许被试在短时间内对视觉目标进行搜索，从而提高搜索效率 (Ngo & Spence, 2012)。第四，以往研究表明，相比于多感觉整合，oddball 刺激吸引注意能更好的解释冰冻效应，当注意力被 oddball 刺激吸引时，会引起主观时间变长(Tse et al., 2004)。本研究中只有实验 2 发现了突然出现的声音冰冻了眼跳，从而通过平均注视时间变长来调节眼动。实验 1 中并没有发现平均注视时间变长，原因可能是受到有空间信息的听觉线索的影响，被试可以通过空间信息来调节眼动，并不需要“冰冻”来搜索视觉目标。

综上四点原因，我们推测 Pip-and-Pop 效应的产生原因有一部分是由多感觉整合所引起的，但也不能完全排除或确定 oddball 吸引注意在其中的作用，所以 Pip-and-Pop 效应的具体产生原因还有待实验进一步探究。

5 结论

视听刺激空间一致性、时间一致性对 Pip-and-Pop 效应具有调节作用。视听刺激空间位置越一致，Pip-and-Pop 效应越大；视听刺激呈现时间越一致，Pip-and-Pop 效应越大。当前研究结果为 Pip-and-Pop 效应产生的原因是多感觉整合提供了证据。此外，在未来的研究中应基于动态视觉搜索范式，设置明确的视听整合条件与 oddball 刺激吸引注意条件进行对比，进一步实验探究 Pip-and-Pop 效应的产生机制。

参考文献

Chamberland, C., Hodgetts, H. M., Vallières, B. R., Vachon, F., & Tremblay, S. (2016). Pip and Pop: When auditory alarms facilitate visual change detection in dynamic settings. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60*(1), 284-288.

Chastain, G., & Cheal, M. (1999). Time course of attention effects with abrupt-onset and offset single- and multiple-element precues. *The American Journal of Psychology*, *112*(3), 411-436.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: *Zeitschriftfur Psychologie*. *217*(3), 108-124.

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, *10*(5), 590-604.

Escera, C., Corral, M. J., & Yago, E. (2002). An electrophysiological and behavioral investigation of involuntary attention towards auditory frequency, duration and intensity changes. *Cognitive Brain Research*, *14*(3), 325-332.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Fister, J. K., Stevenson, R. A., Nidiffer, A. R., Barnett, Z. P., & Wallace, M. T. (2016). Stimulus intensity modulates multisensory temporal processing. *Neuropsychologia*, *88*, 92-100.

Fleming, J. T., Noyce, A. L., & Shinn-Cunningham, B. G. (2020). Audio-visual spatial alignment improves integration in the presence of a competing audio-visual stimulus. *Neuropsychologia*, *146*, 107530.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*(5), 473-490.

Kösem, A., & van Wassenhove, V. (2012). Temporal structure in audiovisual sensory selection. *PLoS One*, *7*(7), e40936.

Knoeferle, K. M., Knoeferle, P., Velasco, C., & Spence, C. (2016). Multisensory brand search: How the meaning of sounds guides consumers' visual attention. *Journal of Experimental Psychology Applied*, *22*(2), 196-210.

Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75-82.

Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory--visual integration. *Behavioural Brain Research*, *121*(1-2), 69-79.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, *16*(3), 468-478.

Lunn, J., Sjoblom, A., Ward, J., Soto-Faraco, S., & Forster, S. (2019). Multisensory enhancement of attention depends on whether you are already paying attention. *Cognition*, *187*, 38-49.

Mishler, A., & Neider, M. (2016). Evidence for the Redundant Signals Effect in Detection of Categorical Targets. *Journal of Vision*, *16*(12), 1024.

Mishler, A. D., & Neider, M. B. (2018). Redundancy gain for categorical targets depends on display configuration and duration. *Visual Cognition*, *26*(6), 393-404.

Ngo, M. K., & Spence, C. (2010a). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention Perception & Psychophysics*, *72*(6), 1654-1665.

Ngo, M. K., & Spence, C. (2010b). Crossmodal facilitation of masked visual target identification. *Attention Perception & Psychophysics*, *72*(7), 1938-1947.

Ngo, M. K., & Spence, C. (2012). Facilitating masked visual target identification with auditory oddball stimuli. *Experimental Brain Research*, *221*(2), 129-136.

Pariyadath, V., & Eagleman, D. (2007). The effect of predictability on subjective duration. *PLoS One*, *2*(11), e1264.

Perrott, D. R., Saberi, K., Brown, K., & Strybel, T. Z. (1990). Auditory psychomotor coordination and visual search performance. *Perception & Psychophysics*, *48*(3), 214-226.

Pluta, S. R., Rowland, B. A., Stanford, T. R., & Stein, B. E. (2011). Alterations to multisensory and unisensory integration by stimulus competition. *Journal Neurophysiology*, *106*(6), 3091-3101.

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *NeuroReport*, *12*(1), 7-10.

Soto-Faraco, S., Kingstone, A., & Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia*, *41*(13), 1847-1862.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy Sciences*, *1296*, 31-49.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *Journal of Cognitive Neuroscience*, *8*(6), 497-506.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*(4), 255-266.

Stevenson, R. A., Fister, J. K., Barnett, Z. P., Nidiffer, A. R., & Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Experimental Brain Research*, *219*(1), 121-137.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sci*ences, *14*(9), 400-410.

Tang, X., Wu, J., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience Biobehavioral Reviews*, *61*, 208-224.

Tsai, Y. Y., & Yeh, S. L. (2013). Freezing effect in tactile perception: sound facilitates tactile identification by enhancing intensity but not duration. *Journal of Experimental Psychology Human Perception & Performance*, *39*(4), 925-932.

Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception & Psychophysics*, *66*(7), 1171-1189.

Van der Brink, R. L., Cohen, M. X., van der Burg, E., Talsma, D., Vissers, M. E., & Slagter, H. A. (2014). Subcortical, modality-specific pathways contribute to multisensory processing in humans. *Cerebral Cortex*, *24*(8), 2169-2177.

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology Human Perception & Performance*, *34*(5), 1053-1065.

Van der Burg, E., Talsma, D., Olivers, C. N., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *Neuroimage*, *55*(3), 1208-1218.

Van der Stoep, N., Van der Stigchel, S., & Nijboer, T. C. (2015). Exogenous spatial attention decreases audiovisual integration. *Attention Perception & Psychophysics*, *77*(2), 464-482.

Van der Stoep, N., Van der Stigchel, S., Nijboer, T. C. W., & Spence, C. (2017). Visually Induced Inhibition of Return Affects the Integration of Auditory and Visual Information. *Perception*, *46*(1), 6-17.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology Human Perception & Performance*, *26*(5), 1583-1590.

Yang, W., Chu, B., Yang, J., Yu, Y., Wu, J., & Yu, S. (2014). Elevated audiovisual temporal interaction in patients with migraine without aura. *Journal of Headache & Pain*, *15*(1), 44.

Zou, H., Müller, H. J., & Shi, Z. (2012). Non-spatial sounds regulate eye movements and enhance visual search. *Journal of Vision*, *12*(5),123-129.

The impact of temporal and spatial consistency of audiovisual stimuli on Pip-and-Pop effect

Abstract

Previous studies usually used the dynamic visual search paradigm to explore the Pip-and-Pop effect, which suggests that the Pip-and-Pop effect is influenced by multiple factors. But the impact of temporal and spatial consistency of audiovisual stimuli on the Pip-and-Pop effect is still controversial. In present study, we applied the eye movement technology to the dynamic visual search paradigm to systematically investigate the impact of temporal and spatial consistency of audiovisual stimuli on the Pip-and-Pop effect. Based on the results of previous studies, we expected that the spatial and temporal consistency of audiovisual modulated the Pip-and-Pop effect. The more consistent the spatial location of audiovisual stimuli are, the greater the Pip-and-Pop effect will be. The more consistent the temporal of audiovisual stimuli are, the greater the Pip-and-Pop effect will be.

Experiment 1 was a 3 (set sizes: 36, 48, 60) × 4 (spatial consistency conditions: ipsilateral, contralateral, bilateral, no sound) within-subjects design, recruited 24 participants(5 males; age range: 19–28 years; mean age: 22.5±2.3 years). The visual search displays consisted of 24, 36, or 48 red (13.9 cd/m^2) or green (46.4 cd/m^2) line segments (0.57°×0.17°) on a black (0.4cd/m^2) background. Auditory stimulus was 1000Hz pure tone (65 dB, 60ms, including 5ms fade-in and 5ms fade-out time), presented through speakers on the left and right sides behind the screen. Each trial started with a drift calibration point. The drift calibration point would not disappear until the participant gazed upon this point. This was followed by a central fixation point of 1000ms, and finally a search screen. The participants were required to find the target line segment and judge whether the target line segment was vertical or horizontal by pressing the key (Z key or M key) as quickly and accurately as possible. Experiment 2 was a 2 (set sizes: 36, 60) × 6 (temporal consistency conditions: -200ms, -100ms, 0ms, 100ms, 200ms, no sound) within-subjects design, recruited 27 participants(5 males; age range: 18–25 years; mean age: 20.7±2.4 years). These temporal consistency conditions represent the tone sounded before (-200, -100ms), simultaneous with (0ms), or after (100, 200ms) the visual target event. The tone could also be absent (no sound). The experimental materials and procedures of Experiment 2 were

identical to those in Experiment1.

Regarding the results in Experiment 1, compared with the baseline condition (no sound), the search response time was the shortest under the condition of ipsilateral, the mean fixation number was the least, and the mean saccade amplitude was the smallest, indicating that the search efficiency was the highest in the ipsilateral condition, the Pip-and-Pop effect was the largest in the ipsilateral condition. Bilateral condition was the second. The contralateral condition showed no significant difference in the above indicators compared with the no sound condition, indicating that no Pip-and-Pop effect was found in the contralateral condition. The results showed that the more consistent the spatial location of audiovisual stimuli were, the greater the Pip-and-Pop effect was. In Experiment 2, compared with the no sound condition, we found that the response time of search was significantly shorter, the mean fixation number in search decreased, the mean saccade amplitude increased, and the mean fixation duration became longer under the conditions of 0ms, -100ms and 100ms. Compared with the no sound condition, conditions -200ms and 200ms showed no significant difference in the above indicators. This indicated that the conditions of 0ms, -100ms and 100ms produced Pip-and-Pop effect, and the Pip-and-Pop effect was the largest under the condition of 0ms, while Pip-and-Pop effect did not find under the condition of -200ms and 200ms. The results showed that the more consistent the temporal of audiovisual stimuli were, the greater the Pip-and-Pop effect was.

In summary, the results showed that the more consistent the audiovisual stimuli were in space and time, the larger the Pip-and-Pop effect was. Therefore, the findings in the present study suggest that the temporal and spatial consistency of audiovisual stimuli modulates the Pip-and-Pop effect. The results provide evidence that the reason for Pip-and-Pop effect is multisensory integration.

Key words Pip-and-Pop effect, dynamic visual search, audiovisual spatial consistency, audiovisual temporal consistency, eye movement technique